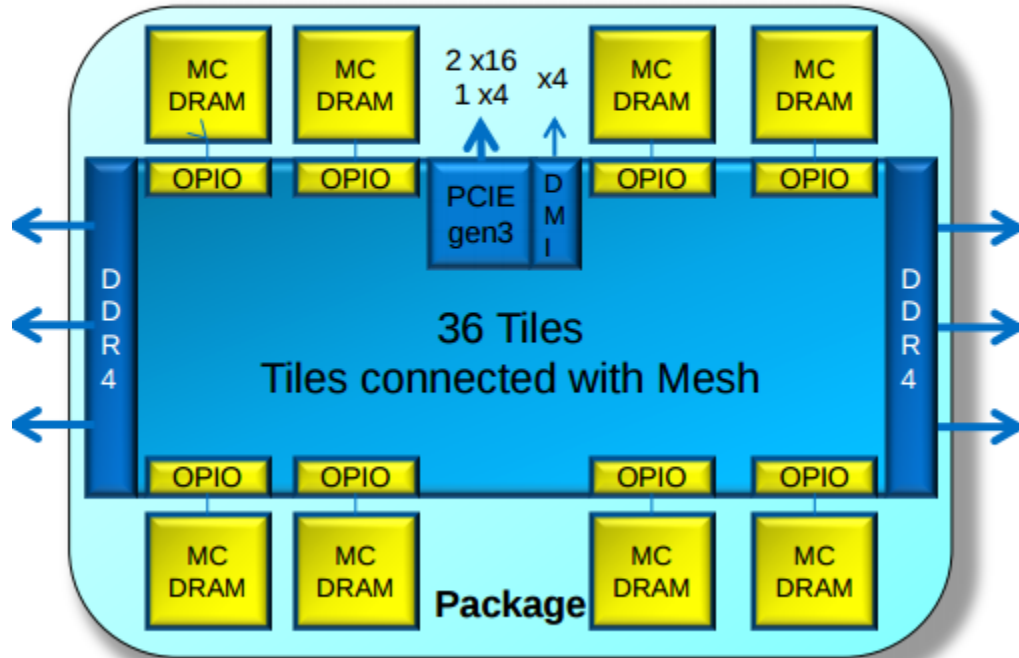


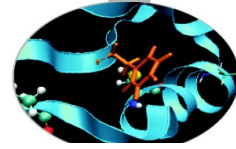
# Introducing the Intel Knights Landing (KNL) architecture

# Knights Landing Overview



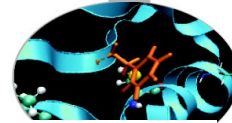
- Stand-alone, Self-boot CPU
- Up to 72 new Silvermont-based cores
- 4 Threads per core. 2 AVX 512 vector units
- Binary Compatible<sup>1</sup> with Intel® Xeon® processor
- 2-dimensional Mesh on-die interconnect
- MCDRAM: On-Package memory: 400+ GB/s of BW<sup>2</sup>
- DDR memory
- Intel® Omni-path Fabric
- 3+ TFlops (DP) peak per package
- ~3x ST performance over KNC

It's not a GPU. It's not an accelerator.  
It's very different from a KNC.



## Many Trailblazing Improvements in KNL

Improvements	What/Why
Self Boot Processor	No PCIe bottleneck
Binary Compatibility with Xeon	Runs all legacy software. No recompilation.
New Core: SLM based	~3x higher ST performance over KNC
Improved Vector density	3+ TFLOPS (DP) peak per chip
AVX 512 ISA	New 512-bit Vector ISA with Masks
Scatter/Gather Engine	Hardware support for gather and scatter
New memory technology: MCDRAM + DDR	Large High Bandwidth Memory → MCDRAM Huge bulk memory → DDR
New on-die interconnect: Mesh	High BW connection between cores and memory



# Intel® AVX Technology



AVX	AVX2
256-bit basic FP	Float16 (IVB 2012)
16 registers	256-bit FP FMA
NDS (and AVX128)	256-bit integer
Improved blend	PERMD
MASKMOV	Gather
Implicit unaligned	

**SNB**

**HSW**

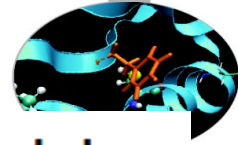
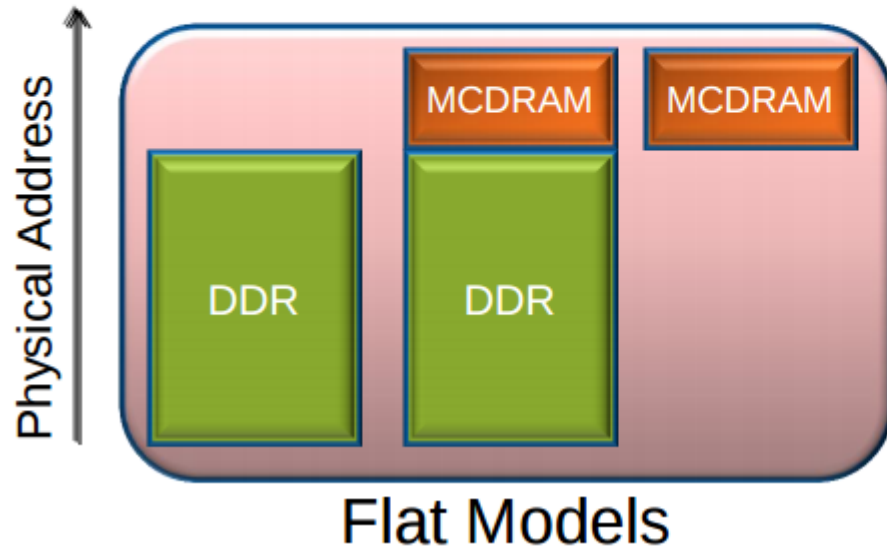
AVX-512
512-bit FP/Integer
32 registers
8 mask registers
Embedded rounding
Embedded broadcast
Scalar/SSE/AVX "promotions"
HPC additions
Gather/Scatter



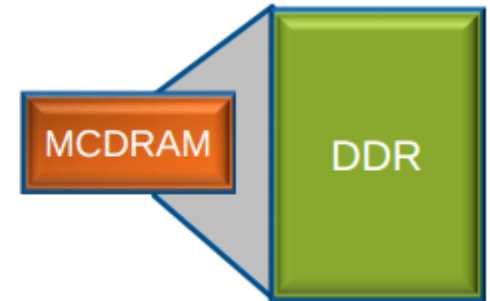
**KNL**

### 3 Memory Modes

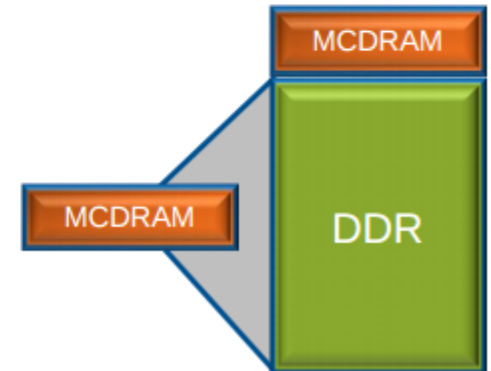
- Mode selected at boot
- MCDRAM-Cache covers all DDR

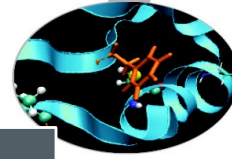


### Cache Model



### Hybrid Model





	DDR Only	MCDRAM as Cache	MCDRAM Only	Flat DDR + MCDRAM	Hybrid
Software Effort	No software changes required			Change allocations for bandwidth-critical data.	
Performance	Not peak performance.		Best performance.		

### MCDRAM exposed as a separate NUMA node

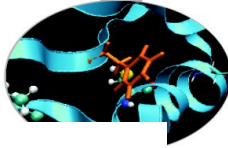


Memory allocated in DDR by default

- Keeps low bandwidth data out of MCDRAM.

Apps explicitly allocate important data in MCDRAM

- **“Fast Malloc”** functions: Built using NUMA allocations functions
- **“Fast Memory”** Compiler Annotation: For use in Fortran.



# Summary

- Knights Landing (KNL) is the first self-boot Intel® Xeon Phi™ processor
- Many improvements for performance and programmability
  - Significant leap in scalar and vector performance
  - Significant increase in memory bandwidth and capacity
  - Binary compatible with Intel® Xeon® processor
- Common programming models between Intel® Xeon® processor and Intel® Xeon Phi™ processor
- KNL offers immense amount of parallelism (both data and thread)
  - Future trend is further increase in parallelism for both Intel® Xeon® processor and Intel® Xeon Phi™ processor
  - Developers need to prepare software to extract full benefits from this trend